How to use Harzing's 'Publish or Perish' software to assess citations – a step-by-step guide

by Blog Admin   July 27, 2011

In his recent blog post on the need for a digital census of academic research, **Patrick Dunleavy** *argued that the 'Publish or Perish' software, developed by Professor Anne-Wil Harzing of Melbourne University and based on Google Scholar data, could provide an exceptionally easy way for academics to record details of their publications and citation instances. An academic with a reasonably distinctive name should be able to compile this report in less than half an hour. Here we present a simple 'how-to' guide to using the software.*

Step 1: Download the software for free from www.harzing.com/pop.htm. The program installs onto your PC.

Step 2: Launch the application from your desktop. When the Harzing PoP screen displays, choose 'Author Impact Analysis' by clicking the tab at top left.

Step 3: The next step has three elements:

(a) Enter the author name you want as initial and surname; in this case we used "J Tinkler". You should always use double quotes around the search terms to ensure accuracy in the results. Choose the most inclusive author name; the more detail you put into the name, the fewer results you will get. "J Tinkler" will garner more results than "Jane Tinkler". Adding extra initials to an author name (e.g. "J R Tinkler") may also reduce the results you get if the author has not consistently published with and/or is not consistently cited by others with both initials.

(b) There are boxes on top right where you can tick which disciplines you want covered. The default is all disciplines, but this will risk generating many 'confuser' results. For most social scientists untick all the fields except **'Business, Administration, Finance, Economics'** and **'Social Sciences, Arts and Humanities'**. However, please note that some publications are unclassified and will not show up if you unclick one or more disciplines. So to be on the safe side, always double-check results with all subject fields ticked, especially if one of your publications seems to be missing.

(c) Click the **Lookup** button on the top right. Compiling the listing will take one or two minutes at most and usually less than 10 seconds. You then see a screen like this, click the image for a larger version.

Step 4:
The main window shows all the works by authors of that name in Google Scholar, listed in order of the number of their citations. You can very easily change the order in which the entries are displayed in the main window by clicking on any of the column headings at the top. Click again to reverse the sort order.

The smaller panel above the main window shows a range of citations statistics, of which the key ones are

- the *total number* of citations

- the *h-index*, where a score of 7 here shows that the author has 7 publications that are each cited at least 7 times, and the author's other publications have no more than 7 citations each and

- the *g-index*, where a score of 19 here shows that the author's top 19 publications taken together have been cited 361 (= g^2) times.

A full account of what each statistic means is provided in the Help file under 'citation metrics'.

Step 5: Cleaning the list is a first priority, to ensure that you are capturing only works by the person that you want. Go through the list of works and in the extreme left column de-click any lines that refer to the wrong author. They will automatically be removed from the statistics, but will still be visible in the main window.

If you see the same confuser names coming up several times, and they are clearly different from the author name you want, then enter them using double quotes into the **Exclude these names** slot near the top of the screen. Now click **Lookup** again and Publish or Perish will rerun the whole analysis with these names kept out. Your list will then look something like the image below.

Step 6: A second clean-up task is called **stacking or merging**. When other authors mis-cite a work, they create duplicate or false titles in Google Scholar in addition to the correct work

Click for larger image
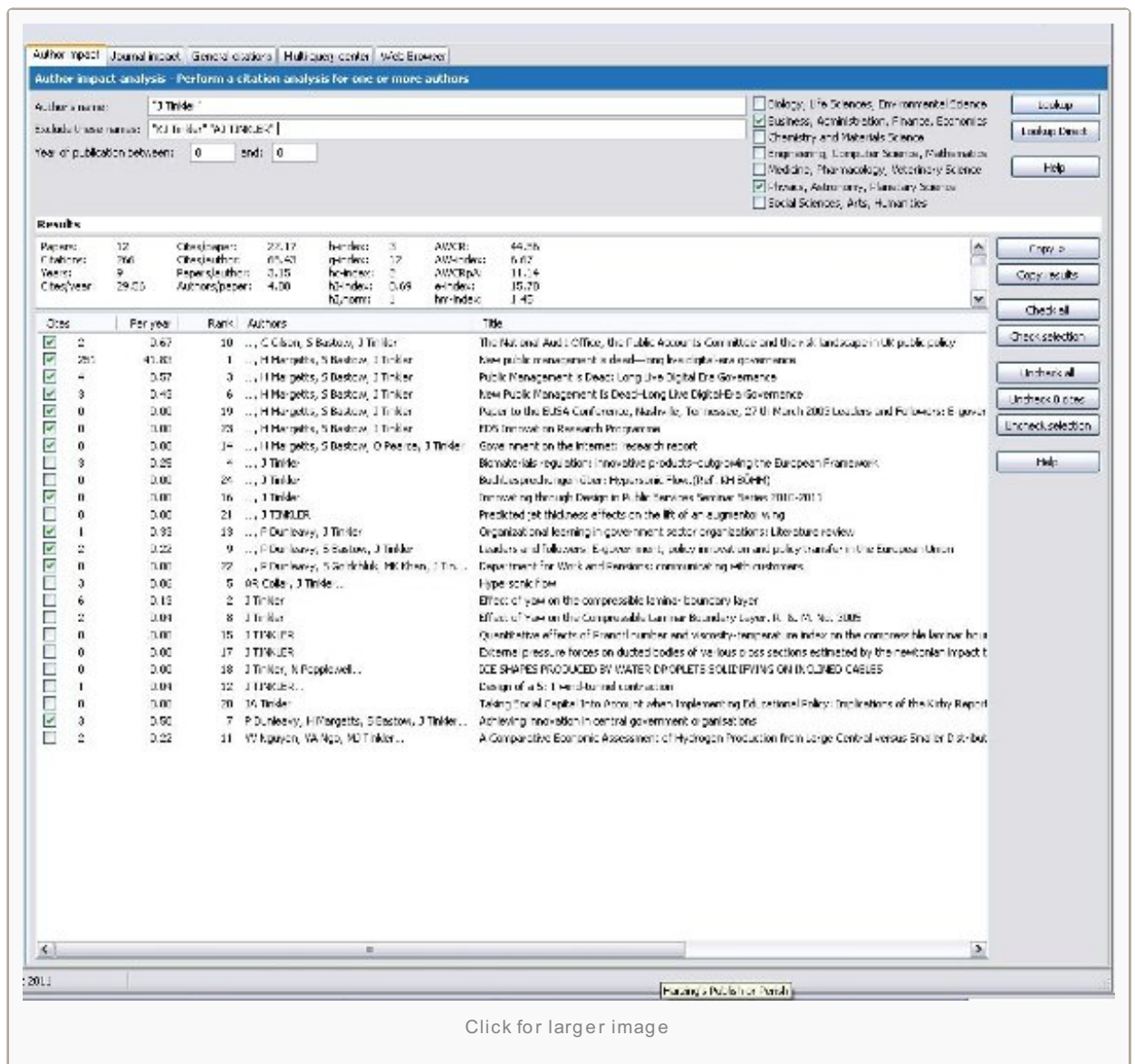
reference, which messes up the statistics. In the list of results, just click on and drag the duplicate entry line and then move it to stack on top of the real entry in the list. This will pool the items for the statistics.

With several duplicates, be careful to choose the most accurate one as the main reference. (If you make a mistake, rightclick the entry on your mouse and click 'Split Citations').

When stacking a large list of publications, it is very useful to click on the 'title' column in the main window, so as to reorder the list in title order for a moment – most mistaken titles are close to the real title and so this will group them next to each other where they can be easily stacked. The other very useful column to click here is the date column, which again will bring items needing stacking into close proximity.

Step 7: Older works by an author (those that get cited at all) tend to accumulate more citations as time goes by. You can get a better comparison of the success of older and newer works by clicking the Per Year column in the main window, which will rearrange the list in order of cites per year. In the statistics all the indices whose names start with AW are age-weighted, and hence also help to discount for simple longevity effects.

Step 8: The list of works can be saved in any order you choose into Excel format (choose 'comma delimited' format) or it can be copied and pasted into Word. (The Word lists may initially look a bit jumbled. If this should happen, they can be quickly clarified by going to the very end of each entry (giving the URL for that work) and clicking forward one space to show the URL in clickable format. Then click return to start the next entry on a new line).

Step 9: But what if you have a *very common author name*, so common that there seems to be no way of

accurately separating your works from those of 'confuser' authors? Those with surnames such as Smith, Dupont or Li will find this is a real problem and a different approach will be required.

Get your CV or other list of your publications and click on the 'General Citations' tab at the top of the Harzing PoP screen. Now put in your author name in the relevant box, and then enter (part of) the title of one of your works in the box labelled 'The Phrase'. You should see the relevant publication clearly in the resulting list. If you get too many results, click the "title words only" box. Go through all your works in turn, copying the lines found each time into Excel before searching for the next item.

You won't get the Publish or Perish statistics this way, but it is still far faster than looking in Google Scholar. With the list in Excel you can also easily compute a few key numbers – total citations, average citations per item and reorder the entries in citation order. From the ordered list it is easy to compute your h-score and g-index following the hints in the *Maximizing the Impacts of the Social Sciences Handbook* (pages 84-9).

Anne-Wil Harzing has also produced The Publish or Perish Book *which provides a lot more detailed guidance on how to use her software, and how to undertake citations analysis in a professional way. It can be purchased as a conventional book by your university library, or in an online version – see* http://www.harzing.com/popbook.htm

Related posts:

1. Why 'Publish or Perish' has the edge over Google Scholar and Scopus when it comes to finding out how your work is used by other academics
2. Your essential 'how-to' guide to using Google Books
3. Google Scholar Citations: a way for academics to compute citation metrics and track them over time
4. Your essential 'how-to' guide to using ISI Web of Knowledge
5. Your essential 'how-to' guide to choosing book titles